Steady States and Constraints in Model Predictive Control

Christopher V. Rao and James B. Rawlings
Dept. of Chemical Engineering, University of Wisconsin, Madison, WI 53706

Studies on the theory of model predictive control include the assumption that the origin is in the interior of the feasible region (that is, the inequality constraints are not active at steady state). The reason for making this assumption is that without it one cannot guarantee feasibility of the control problem on the infinite horizon because of the finite horizon parameterization of the input with an unconstrained linear feedback law. As demonstrated in this article, however, this assumption often does not hold in practice. A strategy for handling inequality constraints active at steady state is presented by projecting the system onto the active constraints under the finite horizon parameterization of the input, as well as an algorithm for constructing the optimal linear feedback law that constrains the system to the active constraints. Feasibility is obtained using output admissible sets. For the steady-state target calculation, we propose an algorithm utilizing exact penalties that treats systems in a unified fashion with more inputs than outputs and vice versa. Assuming the system is detectable, it is proven that the algorithm yields a unique steady-state target.

Introduction

Model Predictive Control (MPC) is an optimization-based strategy that uses a plant model to predict the effect of potential control action on the evolving state of the plant. At each time step, an open-loop optimal control problem is solved and the input profile is injected into the plant until a new measurement becomes available. The updated plant information is used to formulate and solve a new open-loop optimal control problem.

Since MPC is formulated as an optimization problem, inequality constraints are a natural addition to the controller. The ability to handle explicitly input and output constraints may be viewed as one of the major factors for the success of MPC in process control. Operation at constraints is so common that it may be regarded as the rule rather than the exception in chemical process operations. Consider the classic example of temperature control of an exothermic reactor. In order to maximize profit, one may wish to maximize reactor feed rate. At some feed rate, however, the cooling capacity reaches a constraint. As disturbances occur, such as heat exchanger fouling, the feed rate is manipulated to maximize production with some safety margin, while maintaining cooling capacity at its constraint. If a disturbance were to decrease the reactor feed temperature, however, then the cooling rate would be decreased so that the reaction would not extinguish. So, in many practical situations of this type, inputs (cooling rate in this example) are maintained at constraints in the normal steady-state operation. The main objective of this article is to extend the existing MPC theory to handle this important industrial case.

While constraints improve the appeal of MPC as an advanced control strategy, they complicate the implementation of the controller. In addition to the computational burden, constraints necessitate additional safeguards to guarantee that the controller is stabilizing. One method to guarantee nominal stability is to formulate the model predictive controller on an infinite horizon (Keerthi and Gilbert, 1988). Infinite horizon formulations are appealing because, for the nominal case, the predicted open-loop and the achieved closed-loop responses are identical and the effect of tuning parameters is, therefore, more intuitive.

In this article we focus on formulating MPC as an infinite horizon optimal control strategy with a quadratic perfor-

Correspondence concerning this article should be addressed to J. B. Rawlings.

mance criterion. We use the following discrete time model of the plant

$$x_{j+1} = Ax_j + B(u_j + d), \tag{1a}$$

$$y_j = Cx_j + p, \tag{1b}$$

where $x_j \in \mathbb{R}^n$ is the state vector, $u_j \in \mathbb{R}^m$ is the input vector, and $y_j \in \mathbb{R}^q$ is the output vector. $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{q\times m}$ are, respectively, the state transition matrix, the input distribution matrix, and the measurement matrix. The subscript $j \in \mathbb{I}_+$ denotes the discrete time sampling instant. The affine terms $d \in \mathbb{R}^m$ and $p \in \mathbb{R}^q$ serve the purpose of adding integral control. They may be interpreted as modeling the effect of constant disturbances influencing the input and output, respectively. Muske and Rawlings (1993) provide a discussion of how to estimate p and d. Assuming that the state of the plant is perfectly measured, we define MPC as the feedback law $u_j = g(x_j)$ that minimizes

$$\Phi = \frac{1}{2}\sum_{j=0}^{\infty} (y_j - \bar{y})^T Q (y_j - \bar{y}) + (u_j - \bar{u})^T R (u_j - \bar{u}) + \Delta u_j^T S \Delta u_j, \tag{2}$$

where $\Delta u_j \triangleq u_j - u_{j-1}$. The matrices Q, R, and S are assumed to be symmetric positive definite. The vector $\bar{y}$ is the desired output target and $\bar{u}$ is the desired input target, assumed for simplicity to be time-invariant. When the complete state of the plant is not measured, as is almost always the case, the addition of a state estimator is necessary. Since state estimation is beyond the scope of this article, we assume that the control and estimation problems can be separated.

The steady-state aspect of the control problem is to determine appropriate values of (y_{ss}, x_{ss}, u_{ss}) satisfying the following relation

$$x_{ss} = Ax_{ss} + B(u_{ss} + d), \tag{3a}$$

$$y_{ss} = Cx_{ss} + p. \tag{3b}$$

Ideally, $y_{ss} = \bar{y}$ and $u_{ss} = \bar{u}$. However, process limitations and constraints may prevent the system from reaching the desired steady state. The objective of the target calculation is to find the feasible triple (y_{ss}, x_{ss}, u_{ss}) such that y_{ss} and u_{ss} are as close as possible to $\bar{y}$ and $\bar{u}$. We address the target calculation in the next section.

To simplify the analysis and formulation, we transform Eq. 2 using deviation variables to the generic infinite horizon quadratic criterion

$$\Phi = \frac{1}{2}\sum_{j=0}^{\infty} z_j^T Q z_j + \upsilon_j^T R \upsilon_j + \Delta\upsilon_j^T S \Delta\upsilon_j. \tag{4}$$

The original criterion (Eq. 2) can be recovered from Eq. 4 by making the following substitutions

$$z_j \leftarrow y_j - Cx_{ss} - p, \quad w_j \leftarrow x_j - x_{ss}, \quad \upsilon_j \leftarrow u_j - u_{ss}.$$

By using deviation variables, the steady-state and the dynamic elements of the control problem are treated separately, thereby simplifying the overall analysis of the controller.

The dynamic aspect of the control problem is to control (y, x, u) to the steady-state values (y_{ss}, x_{ss}, u_{ss}) in the face of constraints, which may be active at the steady-state operating point. This part of the problem is discussed in the Receding Horizon Regulator section. In particular, we determine the state feedback law $\upsilon_j = \rho(w_j)$ that minimizes Eq. 4. When there are no inequality constraints, the feedback law is the linear quadratic regulator. However, with the addition of inequality constraints, there may not exist an analytic form for $\rho(w_j)$. In such cases where an analytic solution is unavailable, the feedback law is obtained by repetitively solving the open-loop optimal control problem. This strategy allows us to consider only the encountered sequence of measured states rather than the entire state space. For a further discussion, see Mayne (1995).

If we consider only linear constraints on the input, input velocity, and outputs of the form

$$u_{\min} \le Du_k \le u_{\max}, \quad -\Delta_u \le \Delta u_k \le \Delta_u, \quad y_{\min} \le Cx_k \le y_{\max}, \tag{5}$$

where $D \in \mathbb{R}^{n_D \times m}$ and $C \in \mathbb{R}^{n_C \times q}$, we formulate the regulator as the solution to the following infinite horizon optimal control problem

$$\min_{\{w_k, \upsilon_k\}} \Phi(x_j) = \frac{1}{2}\sum_{k=0}^{\infty} z_k^T Q z_k + \upsilon_k^T R \upsilon_k + \Delta\upsilon_k^T S \Delta\upsilon_k, \tag{6}$$

subject to the constraints

$$w_0 = x_j - x_{ss}, \quad \upsilon_{-1} = u_{j-1} - u_{ss},$$

$$w_{k+1} = Aw_k + B\upsilon_k, \quad z_k = Cw_k \tag{7a}$$

$$u_{\min} - u_{ss} \le D\upsilon_k \le u_{\max} - u_{ss}, \quad -\Delta_u \le \Delta\upsilon_k \le \Delta_u, \tag{7b}$$

$$y_{\min} - y_{ss} \le Cw_k \le y_{\max} - y_{ss}. \tag{7c}$$

If we denote

$$\left\{w_{k+1}^*(x_j), \upsilon_k^*(x_j)\right\}_{k=0}^{\infty} = \arg\min \Phi(x_j),$$

then the control law is

$$\rho(x_j) = \upsilon_0^*(x_j).$$

We address the regulation problem in the Receding Horizon Regulator sections.

Combining the solution of the target tracking problem and the constrained regulator, the MPC algorithm is defined as follows:

(1) Obtain an estimate of the state and disturbances $\Rightarrow (x_j, p, d)$

(2) Determine the steady-state target $\Rightarrow (y_{ss}, x_{ss}, u_{ss})$

(3) Solve the regulation problem $\Rightarrow v_j$
(4) Let $u_j = v_j + u_{ss}$
(5) Repeat for $j \leftarrow j+1$

Target Calculation

When the number of the inputs equals the number of outputs, the solution to the unconstrained target problem is obtained using the steady-state gain matrix, assuming such a matrix exists (that is, the system has no integrators). However, for systems with unequal numbers of inputs and outputs, integrators, or inequality constraints, the target calculation is formulated as a mathematical program (Muske and Rawlings, 1993; Muske, 1997). When there are at least as many inputs as outputs, multiple combinations of inputs may yield the desired output target at steady state. For such systems, a mathematical program with a least-squares objective is formulated to determine the best combinations of inputs. When the number of outputs is greater than the number of inputs, situations exist in which no combination of inputs satisfies the output target at steady state. For such cases, we formulate a mathematical program that determines the steady-state output $y_{ss} \neq \bar{y}$ that is closest to $\bar{y}$ in a least-squares sense.

Instead of solving separate problems to establish the target, we prefer to solve one problem for both situations. Through the use of an exact penalty (Fletcher, 1987), we formulate the target tracking problem as a single quadratic program that achieves the output target if possible, and relaxes the problem in a l_1/l_2^2 optimal sense if the target is infeasible. We formulate the soft constraint

$$\begin{aligned} \bar{y} - Cx_{ss} - p &\leq \eta, \\ \bar{y} - Cx_{ss} - p &\geq -\eta, \\ \eta &\geq 0, \end{aligned} \tag{8}$$

by relaxing the constraint $Cx_{ss} = \bar{y}$ using the slack variable η. By suitably penalizing η, we guarantee that the relaxed constraint is binding when it is feasible. We formulate the exact soft constraint by adding an l_1/l_2^2 penalty to the objective function. The l_1/l_2^2 penalty is simply the combination of a linear penalty $q_{ss}^T\eta$ and a quadratic penalty $\eta^T Q_{ss}\eta$, where the elements of q_{ss} are strictly non-negative and Q_{ss} is a symmetric positive definite matrix. By choosing the linear penalty sufficiently large, the soft constraint is guaranteed to be exact. A lower bound on the elements of q_{ss} to ensure that the original hard constraints are satisfied by the solution cannot be calculated explicitly without knowing the solution to the original problem, because the lower bound depends on the optimal Lagrange multipliers for the original problem. In theory, a conservative state-dependent upper bound for these multipliers may be obtained by exploiting the Lipschitz continuity of the quadratic program (Hager, 1979). However, in practice, we rarely need to guarantee that the l_1/l_2^2 penalty is exact. Rather, we use approximate values for q_{ss} obtained by computational experience. In terms of constructing an exact penalty, the quadratic term is superfluous. However, the quadratic term adds an extra degree of freedom for tuning and is necessary to guarantee uniqueness.

We now formulate the target tracking optimization as the following quadratic program

$$\min_{x_{ss},\, u_{ss},\, \eta} \quad \frac{1}{2}\left[\eta^T Q_{ss}\eta + (u_{ss} - \bar{u})^T R_{ss}(u_{ss} - \bar{u})\right] + q_{ss}^T\eta \tag{9}$$

subject to the constraints

$$\begin{bmatrix} I-A & -B & 0 \\ C & 0 & I \\ C & 0 & -I \end{bmatrix} \begin{bmatrix} x_{ss} \\ u_{ss} \\ \eta \end{bmatrix} \begin{Bmatrix} = \\ \geq \\ \leq \end{Bmatrix} \begin{bmatrix} Bd \\ \bar{y} - p \\ \bar{y} - p \end{bmatrix}, \quad \eta \geq 0, \tag{10a}$$

$$u_{\min} \leq Du_{ss} \leq u_{\max}, \quad y_{\min} \leq Cx_{ss} + p \leq y_{\max}, \tag{10b}$$

where R_{ss} and Q_{ss} are assumed to be symmetric positive definite.

Because x_{ss} is not explicitly in the objective function, the question arises as to whether the solution to Eq. 9 is unique. If the feasible region is nonempty, the solution exists because the quadratic program is bounded below on the feasible region (Frank and Wolfe, 1956). If Q_{ss} and R_{ss} are symmetric positive definite, η and u_{ss} are uniquely determined by the solution of the quadratic program. However, without a quadratic penalty on x_{ss}, there is no guarantee that the resulting solution for x_{ss} is unique. Nonuniqueness in the steady-state value of x_{ss} presents potential problems for the controller, because the origin of the regulator is not fixed at each sample time. Consider, for example, a tank where the level is unmeasured (that is, an unobservable integrator). The steady-state solution is to set $u_{ss} = 0$ (that is, balance the flows). However, any level x_{ss}, within bounds, is an optimal alternative. Likewise, at the next time instant, a different level also would be a suitably optimal steady-state target. The resulting closed-loop performance for the system could be erratic, because the controller may constantly adjust the level of the tank, never letting the system settle to a steady state.

In order to avoid such situations, we restrict our discussion to detectable systems, and recommend redesign if a system does not meet this assumption. For detectable systems, the solution to the quadratic program is unique, assuming the feasible region is nonempty. The details of the proof are given in Appendix A. Uniqueness is also guaranteed when only the integrators are observable. For the practitioner, this condition translates into the requirement that all levels are measured. The reason we choose the stronger condition of detectability is that if good control is desired, then the unstable modes of the system should be observable. Detectability is also required to guarantee nominal stability of the regulator.

Empty feasible regions are a result of the inequality constraints (Eq. 10b). Without the inequality constraints (Eq. 10b), the feasible region is nonempty, thereby guaranteeing the existence of a feasible and unique solution under the condition of detectability. For example, the solution $(u_{ss},\ x_{ss},\ \eta) = (-d,\ 0,\ |\bar{y} - p|)$ is feasible. However, the addition of the inequality constraints (Eq. 10b) presents the possibility of infeasibility. Even with well-defined constraints $u_{\min} < u_{\max}$ and $y_{\min} < y_{\max}$, disturbances may render the feasible region empty. Since the constraints on the input usually result from physical limitations such as valve saturation, relaxing only the output constraints is one possibility to circumvent infeasibilities. Assuming that $u_{\min} \le -d \le u_{\max}$, the feasible region is always nonempty. However, we content that the output constraints should not be relaxed in the target calculation. Rather, an infeasible solution, readily determined during the initial phase in the solution of the quadratic program, should be used as an indicator of a process exception. While relaxing the output constraints in the dynamic regulator is common practice (Ricker et al., 1988; Genceli and Nikolaou, 1993; de Oliveira and Biegler, 1994; Zheng and Morari, 1995; Scokaert and Rawlings, 1996, 1997), the output constraint violations are transient. By relaxing output constraints in the target calculation on the other hand, the controller seeks a steady-state target that continuously violates the output constraints. The steady violation indicates that the controller is unable to compensate adequately for the disturbance and, therefore, should indicate a process exception.

Receding Horizon Regulator

Because our implementation of dynamic control in the presence of active steady-state constraints employs an infinite horizon, the solution to infinite horizon problems is briefly discussed.

Infinite horizon optimal control problem

Given the calculated steady state, we formulate the regulator as the following infinite horizon optimal control problem

$$\min_{\{w_k, v_k\}} \Phi(x_j) = \frac{1}{2} \sum_{k=0}^{\infty} w_k^T C^T Q C w_k + v_k^T R v_k + \Delta v_k^T S \Delta v_k, \tag{11}$$

subject to the constraints

$$w_0 = x_j - x_{ss}, \quad v_{-1} = u_{j-1} - u_{ss}, \quad w_{k+1} = A w_k + B v_k, \tag{12a}$$

$$u_{\min} - u_{ss} \le D v_k \le u_{\max} - u_{ss}, \quad -\Delta_u \le \Delta v_k \le \Delta_u, \tag{12b}$$

$$y_{\min} - y_{ss} \le C w_k \le y_{\max} - y_{ss}. \tag{12c}$$

We assume that Q and R are symmetric positive definite matrices. We also assume that the origin $(w_j, v_j) = (0,0)$ is an element of the feasible region $\mathbb{W} \times \mathbb{V}$ ($\mathbb{W} = \{w | y_{\min} - y_{ss} \le Cw \le y_{\max} - y_{ss}\}$, $\mathbb{V} = \{v | u_{\min} \le Dv \le u_{\max}, \ -\Delta_u - u_{ss} \le \Delta v \le \Delta_u - u_{ss}\}$). If the pair (A, B) is stabilizable, the pair $(A, Q^{1/2}C)$ is detectable, and a solution exists to Eqs. 11–12, then $x_j = 0$ is an exponentially stable fixed point of the closed-loop system (Scokaert and Rawlings, 1996).

For unstable state transition matrices, the direct solution of Eqs. 11–12 is ill-conditioned, because the system dynamics are propagated through the unstable A matrix. To improve the conditioning of the optimization, we reparameterize the input as $v_k = L w_k + r_k$, where L is a linear stabilizing feedback gain for (A, B) (Keerthi, 1986; Rossiter et al., 1997). The system model becomes

$$w_{k+1} = (A + BL) w_k + B r_k, \tag{13}$$

where r_k is the new input. By initially specifying a stabilizing, potentially infeasible, trajectory, we can improve the numerical conditioning of the optimization by propagating the system dynamics through the stable $(A + BL)$ matrix.

By expanding Δv_k and substituting in for v_k, we transform Eqs. 11–12 into the following more tractable form

$$\min_{\{w_k, v_k\}} \Phi(x_j) = \frac{1}{2} \sum_{k=0}^{\infty} \left(w_k^T Q w_k + v_k^T R v_k + 2 w_k^T M v_k \right), \tag{14}$$

subject to the following constraints

$$w_0 = x_j, \quad w_{k+1} = A w_k + B v_k, \tag{15a}$$

$$d_{\min} \le D v_k - G w_k \le d_{\max}, \quad y_{\min} - y_{ss} \le C w_k \le y_{\max} - y_{ss}. \tag{15b}$$

The original formulation (Eqs. 11–12) can be recovered from Eqs. 14–15 by making the following substitutions into the second formulation

$$x_j \leftarrow \begin{bmatrix} x_j - x_{ss} \\ u_{j-1} - u_{ss} \end{bmatrix}, \quad w_k \leftarrow \begin{bmatrix} w_k \\ v_{k-1} \end{bmatrix}, \quad v_k \leftarrow r_k,$$

$$A \leftarrow \begin{bmatrix} A + BL & 0 \\ L & 0 \end{bmatrix}, \quad B \leftarrow \begin{bmatrix} B \\ I \end{bmatrix},$$

$$Q \leftarrow \begin{bmatrix} C^T Q C + L^T (R + S) L & -L^T S \\ -SL & S \end{bmatrix}, \quad M \leftarrow \begin{bmatrix} L^T (R + S) \\ -S \end{bmatrix},$$

$$R \leftarrow R + S, \quad D \leftarrow \begin{bmatrix} D \\ I \end{bmatrix}, \quad G \leftarrow \begin{bmatrix} -DL & 0 \\ -L & I \end{bmatrix},$$

$$d_{\max} \leftarrow \begin{bmatrix} u_{\max} - u_{ss} \\ \Delta_u \end{bmatrix}, \quad d_{\min} \leftarrow \begin{bmatrix} u_{\min} - u_{ss} \\ -\Delta_u \end{bmatrix}, \quad C \leftarrow [C \ \ 0].$$

While the formulation (Eqs. 14–15) is theoretically appealing, the solution is intractable in its current form, because it is necessary to consider an infinite number of decision variables. In order to obtain a computationally tractable formulation, we reformulate the optimization in a finite dimensional decision space.

Several authors have considered this problem in various forms. In this article, we concentrate on the constrainted linear quadratic methods proposed in the literature (Keerthi, 1986; Sznaier and Damborg, 1987; Chmielewski and Manousiouthakis, 1996; Scokaert and Rawlings, 1996, 1998). The key concept behind these methods is to recognize that the inequality constraints remain active only for a finite number of sample steps along the prediction horizon. We demonstrate informally this concept as follows: if we assume that there exists a feasible solution to Eqs. 14 and 15, then the state and input trajectories $\{w_k, v_k\}_{k=0}^{\infty}$ approach the origin exponentially. Furthermore, if we assume the origin is contained within the interior of the feasible region $\mathbb{W} \times \mathbb{V}$ (we address the case where the origin lies on the boundary of the feasible region in the next section), then there exists a positively invariant convex set (Gilbert and Tan, 1991)

$$\mathcal{O}_\infty = \left\{ w | (A + BK)^j w \in \mathbb{W}_K, \quad \forall j \geq 0 \right\} \tag{16}$$

such that the optimal unconstrained feedback law $v = Kw$ is feasible for all future time. The set $\mathbb{W}_K$ is the feasible region projected onto the state space by the linear control K (that is, $\mathbb{W}_K = \{w | (w, Kw) \in \mathbb{W} \times \mathbb{V}\}$). Because the state and input trajectories approach the origin exponentially, there exists a finite N^* such that the state trajectory $\{w_k\}_{k=N^*}^{\infty}$ is contained in $\mathcal{O}_\infty$.

In order to guarantee that the inequality constraints (Eq. 15b) are satisfied on the infinite horizon, N^* must be chosen such that $w_{N^*} \in \mathcal{O}_\infty$. Since the value of N^* depends on x_j, we need to account for the variable decision horizon length in the optimization. We formulate the variable horizon length regulator as the following optimization

$$\min_{\{w_k, v_k, N\}} \Phi(x_j) = \frac{1}{2} \sum_{k=0}^{N-1} \left(w_k^T Q w_k + v_k^T R v_k + 2 w_k^T M v_k \right) + \frac{1}{2} w_N^T \Pi w_N, \tag{17}$$

subject to the constraints

$$w_0 = x_j, \quad w_{k+1} = A w_k + B v_k, \tag{18a}$$

$$d_{\min} \leq D v_k - G w_k \leq d_{\max}, \quad y_{\min} - y_{ss} \leq C w_k \leq y_{\max} - y_{ss}, \tag{18b}$$

$$w_N \in \mathcal{O}_\infty. \tag{18c}$$

The cost to go Π is determined from the discrete-time algebraic Riccati equation

$$\Pi = A^T \Pi A + Q - (A^T \Pi B + M)(R + B^T \Pi B)^{-1}(B^T \Pi A + M^T), \tag{19}$$

for which many reliable solution algorithms exist. The variable horizon formulation is similar to the dual-mode receding horizon controller (Michalska and Mayne, 1993) for a nonlinear system with the linear quadratic regulator chosen as the stabilizing linear controller.

While the problem (Eqs. 17–18) is formulated on a finite horizon, the solution cannot, in general, be obtained in real time since the problem is a mixed-integer program. Rather than try to directly solve Eqs. 17–18, we address the problem of determining N^* from a variety of semi-implicit schemes, while maintaining the quadratic programming structure in the subsequent optimizations.

Gilbert and Tan (1991) show that there exist a finite number t^* such that $\mathcal{O}_{t^*}$ is equivalent to the maximal $\mathcal{O}_\infty$, where

$$\mathcal{O}_t = \left\{ w | (A + BK)^j w \in \mathbb{W}_K, \text{ for } j = 0, \ldots, t \right\}. \tag{20}$$

They also present an algorithm for determining t^* that is formulated efficiently as a finite number of linear programs. Their method provides an easy check whether, for a fixed N, the solution to Eqs. 17–18 is feasible (that is, $w_N \in \mathcal{O}_\infty$). The check consists of determining whether state and input trajectories generated by unconstrained control law $v_k = K w_k$ from the initial condition w_N are feasible with respect to inequality constraints for t^* time steps in the future. If the check fails, then the optimization (Eqs. 17–18) needs to be resolved with a longer control horizon $N' > N$ since $w_N \notin \mathcal{O}_\infty$. The process is repeated until $w_{N'} \in \mathcal{O}_\infty$.

When the set of initial conditions $\{w_0\}$ is compact, Chmielewski and Manousiouthakis (1996) present a method for calculating an upper bound $\overline{N}$ on N^* using boundary arguments on the optimal cost function Φ^*. Given a set $\mathbb{P} = \{x^1, \ldots, x^m\}$ of initial conditions, the optimal cost function $\Phi^*(x)$ is a convex function defined on the convex hull (co) of $\mathbb{P}$. An upper bound $\overline{\Phi}(x)$ on the optimal cost $\Phi^*(x)$ for $x \in \mathrm{co}(\mathbb{P})$ is obtained by the corresponding convex combinations of optimal cost functions $\Phi^*(x^j)$ for $x^j \in \mathbb{P}$. The upper bound on N^* is obtained by recognizing that the state trajectory w_j only remains outside of $\mathcal{O}_\infty$ for a finite number of stages. A lower bound q on the cost of $w_j^T Q w_j$ can be generated for $x_j \notin \mathcal{O}_\infty$ [see Chmielewski and Manousiouthakis (1996) for explicit details]. It then follows that $N^* \leq \overline{\Phi}(x)/q$. Further refinement of the upper bound can be obtained by including the terminal stage penalty Π in the analysis.

When a bound on the initial conditions w_0 is known *a priori*, the Chmielewski and Manousiouthakis method is appealing, because one need not iteratively determine N^* on-line. However, generating this bound *a priori* requires significant process knowledge. Changing operating conditions and disturbances may lead to initial conditions that violate any previously specified bound. In such cases, we again need to determine N^* on-line. Furthermore, the decision of how to construct the basis for $\mathbb{P}$ is complicated, since the number of points increases exponentially in higher dimensions. Even when a bound is available and a logical basis is constructed, the upper bounds are often conservative, as demonstrated in the following example.

Example 1. Comparison of On-line and Off-line Determination of N^*. Consider the regulation of the following double integrator system

$$\dot{x}_1 = x_2, \tag{21a}$$

$$\dot{x}_2 = u, \tag{21b}$$

sampled at a frequency of 10 Hz with $y = x_1$ and the input constraint $|u| \leq 1$. For $Q = 1$, $R = 1$, $S = 0$, and the initial condition, $x_0 = [1\ 1]^T$. For this initial condition, $N^* = 13$ was required to guarantee that the constraints are satisfied on the infinite horizon.

The Chmielewski and Manousiouthakis method generates a least upper bound of 361 for N^*. This value was determined using the true infinite horizon cost for $x_0 = [1\ 1]^T$. In practice, only an upper bound on the cost is available for the infinite horizon cost, so the upper bound on N^* is often greater than the least upper bound for N^*. We can compare these results with the repetitive strategy where N is increased until $w_N \in \mathcal{O}_\infty$. Since there exist algorithms whose computational cost is $O(N)$ (Rao et al., 1998), we can expect that the computational cost is approximately a linear function of N. If $N=1$ initially and the control horizon is increased by unit steps, then the total computational cost is approximately $91\times C$, where C is the computational cost required solve the optimization for $N=1$. If we increase the horizon geometrically with a factor of 2, as advocated by Scokaert and Rawlings (1998), then the total computational cost is approximately $31\times C$. In practice, larger initial values of N are used. A good heuristic is to choose initially $N=t^*$. For this example, $t^*=15$. As the example demonstrates, the on-line determination is significantly less computationally expensive than the off-line determination. Furthermore, for the on-line determination, we can bound the computational cost by $4N^* \times C$, for this example $52\times C$, when we increase the horizon geometrically with a factor of 2 (Scokaert and Rawlings, 1998). With the off-line determination, we have no bounds on the computational cost (other than it is finite), and, as the example demonstrates, a computational effort an order of magnitude greater than required is possible. Therefore, we suggest the use of the iterative, on-line determination for N^*.

Boundary solutions and suboptimal approximations

All articles on constrained linear MPC include the assumption that the origin lies in the interior of the feasible region (Keerthi and Gilbert, 1988; Sznaier and Damborg, 1987; Rawlings and Muske, 1993; Chmielewski and Manousiouthakis, 1996; Scokaert and Rawlings, 1996). However, as the Target Calculation section indicates, this assumption is often violated. In practice, one often encounters situations in which a valve saturates or a control variable rides at a performance constraint during steady-state operation. In these situations, the origin is on the boundary of the feasible region. Table 1 lists all of the examples that are discussed in this article and summarizes the main points illustrated with each example. Consider the following example.

Example 2. Saturating Inputs at Steady State. Prett and Morari (1987) presented the following model

$$G(s)=\begin{bmatrix} \dfrac{4.05e^{-27s}}{50s+1} & \dfrac{1.77e^{-28s}}{60s+1} & \dfrac{5.88e^{-27s}}{50s+1} \\ \dfrac{5.39e^{-18s}}{50s+1} & \dfrac{5.72e^{-14s}}{60s+1} & \dfrac{6.90e^{-15s}}{40s+1} \\ \dfrac{4.38e^{-20s}}{33s+1} & \dfrac{4.42e^{-22s}}{44s+1} & \dfrac{7.20}{19s+1} \end{bmatrix} \tag{22}$$

Table 1. Brief Synopsis of Examples

Example	Description
1	A comparison of methods for calculating N^* in solving infinite horizon problems
2	Steady-state inputs on boundary for Shell problem
3	Steady-state outputs on boundary for Furnace problem
4	Example of system whose input never settles on constraint or remains in interior of feasible region
5	Simple numerical example of input constrained regulator
5	Dynamic response of Shell problem subject to set point change
6	Dynamic response of Shell problem subject to disturbance
7	Example of output constrained regulator: endpoint constraint necessary
8	Example of output constrained regulator: boundary solution

for a heavy oil fractionator as the benchmark process for the Shell standard control problem. The three inputs of the process represent the product draw rate from the top of the column (u_1), the product draw rate from the side of the column (u_2), and the reflux heat duty for the bottom of the column (u_3). The three outputs of the process represent the draw composition (y_1) from the top of the column, the draw composition (y_2) from the side of the column, and the reflux temperature at the bottom of the column (y_3). Prett and Garcia also present the following disturbance model

$$G_d(s)=\begin{bmatrix} \dfrac{1.20e^{-27s}}{45s+1} & \dfrac{1.44e^{-27s}}{40s+1} \\ \dfrac{1.52e^{-15s}}{25s+1} & \dfrac{1.83e^{-15s}}{20s+1} \\ \dfrac{1.14}{27s+1} & \dfrac{1.26}{32s+1} \end{bmatrix} \tag{23}$$

for the heavy oil fractionator. The two disturbances are the reflux heat duty for the intermediate section of the column (d_1) and the reflux heat duty for the top of the column (d_2). Both models were sampled with a period of 4 min.

The inputs are constrained between -0.5 and 0.5. An input velocity constraint of 0.20 is also imposed. In addition to constraints on the inputs, the outputs are constrained between -0.5 and 0.5. The following tuning parameters were chosen: $Q(Q_{ss})=I$ and $R(R_{ss})=I$.

Since the origin is shifted by the steady-state target calculation, output target changes and measured disturbances may force the origin to lie on the boundary of the feasible region. An example of an output target change that causes the inputs to saturate at steady state is

$$\bar{y}=\begin{bmatrix} 0.3 \\ 0.3 \\ -0.3 \end{bmatrix} \Rightarrow u_{ss}=\begin{bmatrix} 0.5 \\ -0.1 \\ -0.26 \end{bmatrix}, \quad y_{ss}=\begin{bmatrix} 0.3 \\ 0.3 \\ -0.15 \end{bmatrix}. \tag{24}$$

Note that since the input (u_1) saturates, the system is unable to attain the desired target (that is, $\bar{y}\neq y_{ss}$). Figure 1 illustrates how the input constraints constrain the attainable region of the output space. Likewise, an example of a step disturbance d_{step} that causes the inputs to saturate at steady state is

$$d_{\text{step}}=\begin{bmatrix} 0.6 \\ 0.5 \end{bmatrix} \Rightarrow u_{ss}=\begin{bmatrix} -0.5 \\ 0.04 \\ 0.09 \end{bmatrix}. \tag{25}$$

Steady-state outputs at performance constraints are a consequence of choosing an output target at ± 0.5 or choosing an infeasible target.

Example 3. Constrained Outputs at Steady State. Consider the control of a furnace depicted in Figure 2, where the objective is to preheat the feed to a desired output temperature. The input variable is the fuel-gas flow rate. In addition to the input constraint caused by valve saturation, there is a maximum limit for the furnace temperature in order to prevent the furnace tubes from melting. While temporary violations of the constraint are tolerable, long-term violations are

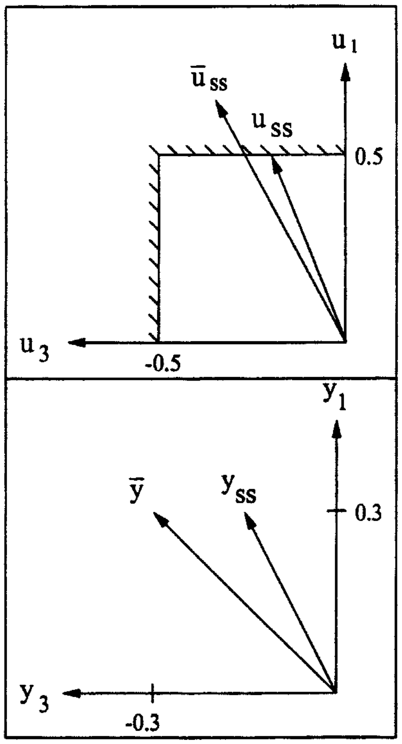

Figure 1. Limiting effect of the input constraints on the ability to attain the desired output target for Example 2.

The vector $\bar{u}_{ss}$ denotes the unconstrained input required to achieve the target $\bar{y}$.

not. If we assume that the heating tube temperature and the heat transfer are linear functions of the fuel-gas flow rate, then a simplified, steady-state energy balance neglecting heat loss yields the following dimensionless model for the furnace system

$$y_1 = \alpha u, \tag{26a}$$

$$y_2 = \beta u + d, \tag{26b}$$

in which y_1 is the heating tube temperature, y_2 is the outlet temperature, d is the inlet temperature, and u is the fuel-gas

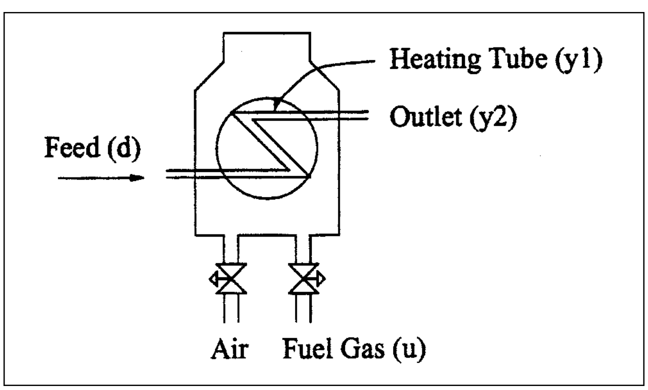

Figure 2. Preheater furnace.

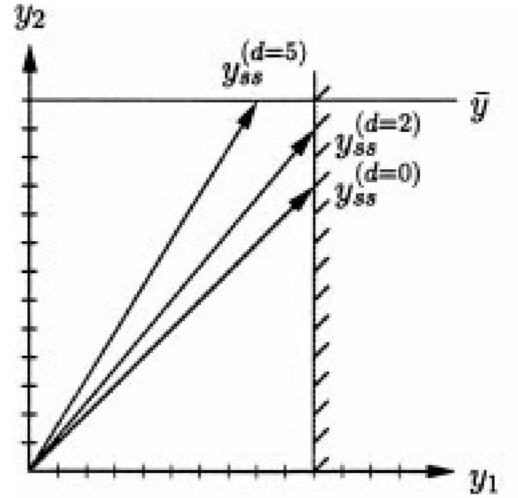

Figure 3. Inability of the outlet temperature to reach the desired steady-state target due to the tube temperature constraint in Example 3.

flow rate. For simplicity, we scale the variable such that $\alpha = \beta = 1$. Assume nominal conditions are $u = 8$, $d = 5$, and $\bar{y} = [8\ 13]^T$, and the maximum limit for the furnace temperature is $y_1 \leq 10$. Suppose that there is an upstream disturbance that causes the inlet temperature to drop to $d = 2$. To compensate for the disturbance, the fuel-gas flow rate would have to increase to 11 for the output temperature (y_2) to remain at its target. However, the furnace temperature constraint would allow the flow rate to increase only to 10. The resulting steady-state output obtained with $Q_{ss} = 1$, $R_{ss} = 1$, $q_{ss} = 100$, and $\bar{u} = 0$ is $y_{ss} = [10\ 12]^T$, which lies on the boundary of the feasible region. Figure 3 depicts the effect of the disturbance on the attainable region of the output space.

These situations complicate the formulation of the infinite horizon optimization, because $\mathcal{O}_\infty$ does not exist for all such systems controlled with the unconstrained optimal feedback regulator. Figure 4 displays some of the potential input and state trajectory characteristics that are possible when the origin lies on the boundary of the feasible region.

An example of a system that displays the first characteristic is a stable first-order system with initial conditions in the interior of the feasible region. An example of a system displaying the second characteristic is given in the following example.

Example 4. A System Where the Control does not Become Permanently Active or Inactive on the Constraint. Consider the following system

$$w_{k+1} = \begin{bmatrix} 0.5477 & 0.8208 & 0 \\ -0.8208 & 0.5067 & 0 \\ 0 & 0 & 0.8 \end{bmatrix} w_k, + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} v_k, \tag{27a}$$

$$y_k = [1 \quad 0 \quad 1] w_k, \tag{27b}$$

subject to the input constraint $v_k \leq 0$. Figure 5 details the optimal input profile subject to the initial disturbance $w_0 = [3\ 3\ 0]^T$ with tuning parameters $Q = 1$ and $R = 1$. While the input trajectory converges toward the origin (see Figure 5), nu-

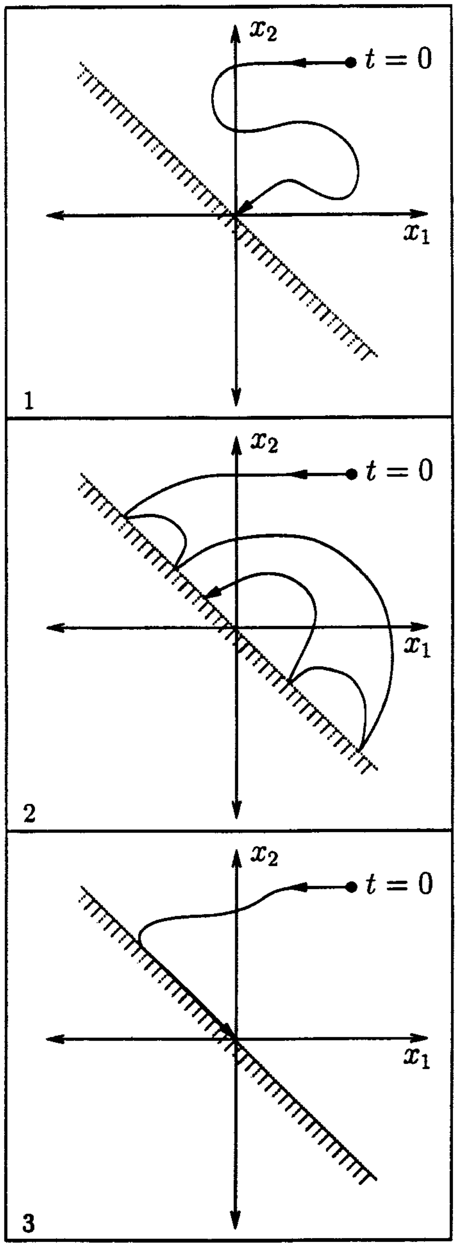

Figure 4. Some potential input and state trajectories when the origin is on the boundary of the feasible region.

To prevent problems associated with trajectories 1 and 2, the proposed algorithm forces the closed-loop response to adhere to a trajectory similar to 3.

merical calculations indicate that the input does not become active on the constraint or stay strictly in the interior of the feasible region.

Examples of systems displaying the third characteristic are given in Examples 7 and 8. While the first trajectory offers the possibility of constructing $\mathcal{O}_\infty$, for the second and third trajectory we are unable to construct $\mathcal{O}_\infty$ with a finite number of inequality constraints because the constraints remains active for infinite time. Each of the possibilities could be handled individually. However, the task of segregating their behavior *a priori* is difficult. To circumvent this problem, we

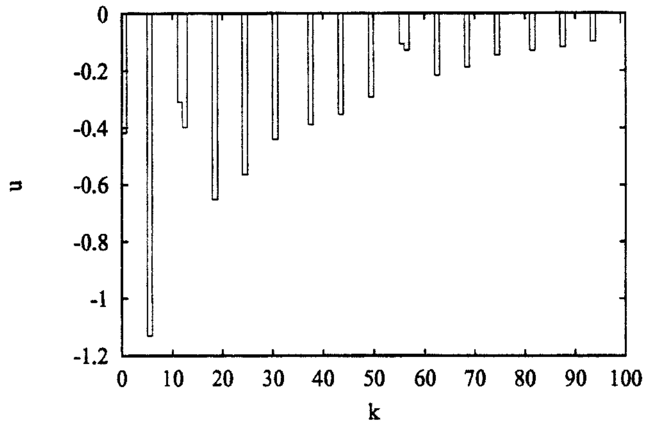

Figure 5. Input trajectory for Example 4.

approximate Eqs. 17–18 by restricting the evolution of the input and state trajectories generated by the linear control law to the null space of the active constraints at the origin. This suboptimal strategy coincides with forcing the state and input trajectories to adhere to the third path depicted in Figure 4. By forcing the trajectory of system onto the constraints active at the origin, we guarantee the existence of an $\mathcal{O}_\infty$, because the invariant set needs to account only for the constraints inactive at the origin. The constraints active at the origin are feasible by construction.

We accomplish the boundary approximation by constructing the optimal linear feedback law $\overline{K}$ that constrains evolution of the closed-loop system to constraints active at the origin. The set $\mathcal{O}_\infty$ is constructed as before with the following differences: the new linear control law is $\overline{K}$ and the inequality constraints active at the origin are discarded. The infinite horizon regulator is constructed as the solution of Eqs. 17–18 with the new $\mathcal{O}_\infty$ and the cost to go $\overline{\Pi}$ associated with $\overline{K}$. We treat the situation of input and state constraints separately.

Active Input Constraints. We recast the problem for handling input constraints active at the origin as finding the linear feedback controller that minimizes the infinite horizon quadratic objective (Eq. 14) subject to the equality constraints (it is not necessary to account for $\overline{G}$ because the velocity constraints are not active at steady state)

$$w_{k+1} = Aw_k + Bv_k, \tag{28a}$$

$$\overline{D}v_k = 0, \tag{28b}$$

where the overbar denotes the subset of the inequality constraints $\underline{d}_{\min} \le Dv_k \le d_{\max}$ that are active at the origin. The matrix $\overline{D} \in \mathbb{R}^{n_{\overline{D}} \times m}$, where $n_{\overline{D}}$ is equal to the number of inputs with constraints active at the origin. The state dependence of the input constraints due to Eq. 13 can be reformulated solely in terms of the input v_k by removing the parameterization $v_k = Lw_k + r_k$ for $k \ge N$. We derive the linear optimal controller as follows. We first define the operator

$$\mathcal{K}: (A, B, Q, R, M) \rightarrow (K, \Pi), \tag{29}$$

where K is the linear gain for the optimal unconstrained regulator and Π is the solution to the associated Riccati equation. If we let $\mathfrak{N}_{\overline{D}}$ be an orthonormal basis for the null space of $\overline{D}$, then $p_k^{\overline{D}} = \mathfrak{N}_{\overline{D}}^T v_k$ represents the input projected to the null space $\mathfrak{N}_{\overline{D}}$. Because the equality constraint is feasible for all $p_k^{\overline{D}}$, we can substitute for v_k in the state equation and the objective function yielding the following expression

$$\left(K_{\overline{D}}, \overline{\Pi}\right) = \mathcal{K}\left(A, B\mathfrak{N}_{\overline{D}}, Q, \mathfrak{N}_{\overline{D}}^T R\, \mathfrak{N}_{\overline{D}}, M\mathfrak{N}_{\overline{D}}\right) \tag{30}$$

for the solution of the constrained feedback law. If $(A, B\mathfrak{N}_{\overline{D}})$ is stabilizable, then

$$\overline{K} \triangleq \mathfrak{N}_{\overline{D}} K_{\overline{D}}. \tag{31}$$

If $(A, B\mathfrak{N}_{\overline{D}})$ is not stabilizable, we need to zero the modes of the system that are both uncontrollable and unstable at $k = N^*$ in order to guarantee nominal stability. By first performing a Kalman decomposition to construct a basis for the uncontrollable subspace, a basis for the corresponding uncontrollable and unstable modes is constructed using either a Jordan or a Schur decomposition. We remark that if $\overline{D}$ is full rank, then it is necessary to zero the inputs at the end of the control horizon. The regulator then reduces to the one discussed by Rawlings and Muske (1993). In the following two examples we show the closed-loop response of the heavy oil fractionator with the output target change and disturbance described in Example 2.

Example 5. Heavy Oil Fractionator: Closed-Loop Response Subject to Output Target Change. Consider the closed-loop response of the heavy oil fractionator described in Example 2 subject to the output target change described in Eq. 24. Figure 6 shows the closed-loop response subject to the output target change. As discussed in Example 2, the input constraints prevent the system from attaining the desired target $\bar{y}$ at steady state. Instead the controller seeks a target that causes the top draw (u_1) to saturate at its upper limit. Figure 6 shows that, after initially saturating, the top draw asymptotically approaches its upper limit as the closed-loop system settles at its specified steady-state values. While the open-loop trajectory of the controller specifies that the top draw saturates at $k = N^*$, the receding horizon aspect of the regulator allows for an asymptotic approach. This feedback effect diminishes the performance degradation due to the boundary projection in the actual closed loop.

Example 6. Heavy Oil Fractionator: Closed-Loop Response Subject to a Disturbance. Consider the closed-loop response of the heavy oil fractionator described in Example 2 subject to the disturbance described in Eq. 25. An output disturbance model was used to detect the disturbance (Muske and Rawlings, 1993). Figure 7 displays the closed-loop response subject to the disturbance. To reject the disturbance, the controller seeks an input target that causes the top draw (u_1) to saturate at its lower limit. Once again, Figure 7 shows that the top draw asymptotically approaches its lower limit as the closed-loop system settles at its specified steady state. In addition, the disturbance causes the output constraints to become infeasible. In order to handle the output infeasibilities in the regulator, the constraints were relaxed using a l_1/l_2^2 exact penalty in the manner described by Scokaert and Rawlings (1996). The output constraints were relaxed with an l_1 penalty of $z = 1{,}000 * e$ and a l_2^2 penalty of $Z = I$, where $e = [1 \ \cdots \ 1]^T$.

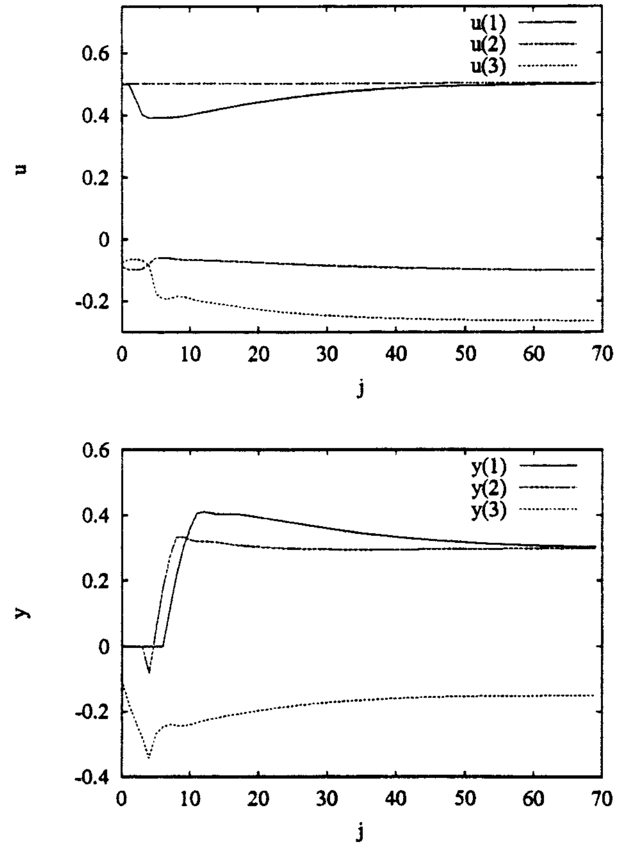

Figure 6. Closed-loop response for Example 5.

Active State Constraints. In an analogous manner, the problem of handling state constraints can be reformulated as finding the linear feedback controller that minimizes the infinite horizon quadratic objective (Eq. 14) subject to the constraints

$$w_{k+1} = Aw_k + Bv_k, \tag{32a}$$

$$\overline{C}w_k = 0. \tag{32b}$$

Unlike the previous situation, there does not always exist a linear feedback regulator that satisfies the state constraints for all k. For such a regulator to exist, we require that the null space of $\overline{C}$ is (A, B) invariant. The definition of (A, B) invariance along with the sufficient conditions for the existence of a regulator is given in Appendix B. The condition of (A, B) invariance essentially requires that there are enough degrees of freedom in the input to constrain the evolution of the system to a particular subspace. A system whose uncontrollable modes are observable in the null space of $\overline{C}$ is not (A, B) invariant if the associated basis for the uncontrollable modes is not contained completely in the null space of $\overline{C}$. Assuming that the null space of $\overline{C}$ is (A, B) invariant, then

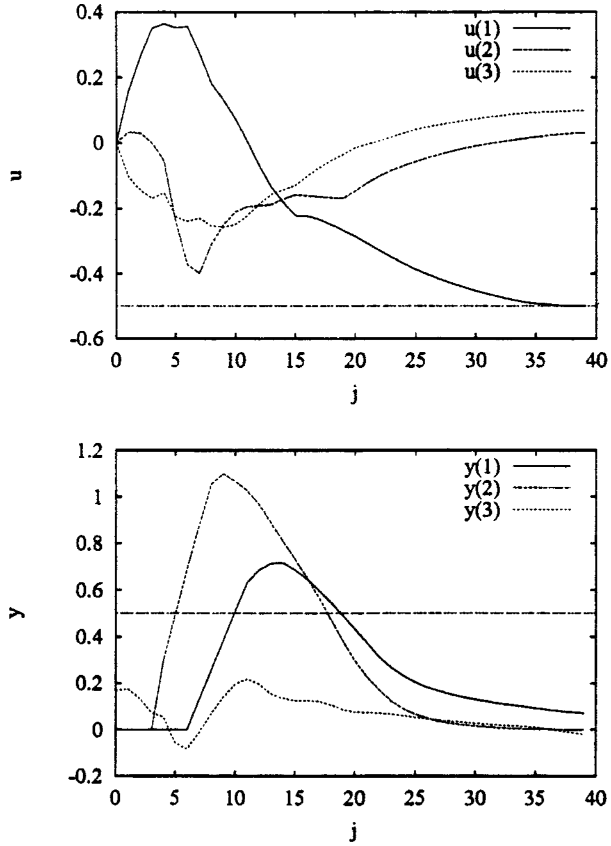

Figure 7. Closed-loop response for Example 6.

there exists a linear feedback law

$$\upsilon_k = L_{\overline{C}} w_k + \mathfrak{N}_{B\overline{C}} p_k^{\overline{C}} \tag{33}$$

that constrains the system to the null-space of $\overline{C}$ for all $p_k^{\overline{C}}$. The details of the construction are given in Appendix B. We obtain the optimal feedback law from the following expression

$$\left(K_{\overline{C}}, \overline{\Pi}\right) = \mathcal{K}\Big(\left(A + BL_{\overline{C}}\right), B\mathfrak{N}_{\overline{C}B}, Q + L_{\overline{C}}^T R L_{\overline{C}} + 2ML_{\overline{C}},$$

$$\mathfrak{N}_{\overline{C}B}^T R\, \mathfrak{N}_{\overline{C}B}, M\mathfrak{N}_{\overline{C}B} + L_{\overline{C}}^T R\, \mathfrak{N}_{\overline{C}B}\Big) \tag{34}$$

by substituting in for υ_k with the feedback law (Eq. 33). If $(A + BL_{\overline{C}},\ B\mathfrak{N}_{CB})$ is stabilizable, then

$$\overline{K} \triangleq \mathfrak{N}_{\overline{C}B} K_{\overline{C}}. \tag{35}$$

Otherwise, it is necessary to zero the unstable and uncontrollable modes at $k = N^*$ in an analogous manner to the input constrained regulator. The boundary approximation to Eqs. 17–18 is obtained by adding the constraint

$$\overline{C} w_N = 0, \tag{36}$$

and calculating $\overline{\Pi}$ using Eq. 34. In the following two examples, we illustrate that a system must possess excess degrees of freedom in the input for a stabilizing boundary approximation.

Example 7. Output Constrained Regulator with No Excess Degrees of Freedom. Consider the regulation of the following nonminimum phase system

$$y(s) = \frac{s-3}{3s^2 + 4s + 2} u(s) \tag{37}$$

sampled at a frequency of 10 Hz subject to the constraint $y_k \geq 0$ and the tuning parameters $Q = 1$ and $R = 1$. One state space realization for this system in discrete time is

$$A = \begin{bmatrix} 0.9968 & 0.0935 \\ -0.0623 & 0.8721 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0048 \\ 0.0935 \end{bmatrix}, \tag{38a}$$

$$C = [-1.0000 \quad 0.3333]. \tag{38b}$$

The system is (A, B) invariant with respect to the null space of C. However, the closed-loop system with the invariant feedback law

$$K_{\overline{C}} = [38.5608 \quad -7.4723] \tag{39}$$

is unstable. Since there are no additional degrees of freedom, it is necessary to enforce the endpoint constraint $w_N = 0$ for the boundary approximation. Figure 8 shows a comparison of

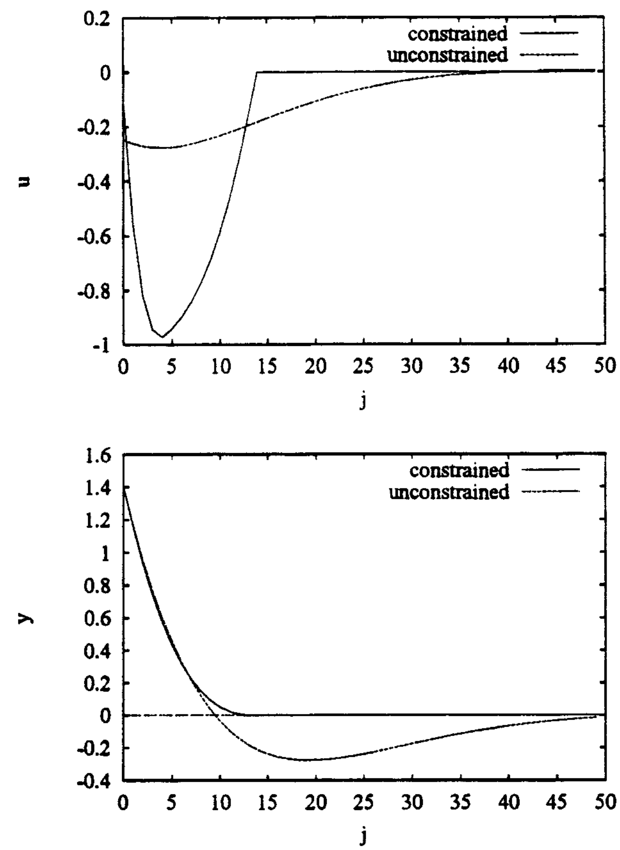

Figure 8. Comparison of closed-loop responses for Example 7.

the closed-loop response for the constrained regulator with $N=10$ and the unconstrained regulator with an initial state disturbance of $[-1,2]^T$. While the output response for the constrained regulator displayed in Figure 8 appears better than the unconstrained output response, notice that the input action is far more aggressive for the constrained regulator.

Example 8. Output Constrained Regulator with Excess Degrees of Freedom. Reconsider the regulation of the system in Example 7 with an additional input

$$y(s)=\frac{s-3}{3s^2+4s+2}u_1(s)+\frac{2}{3s^2+4s+2}u_2(s) \tag{40}$$

sampled at a frequency of 10 Hz subject to the constraints $|u_k|\leq 6$ and $y_k\geq 0$ and the tuning parameters $Q=1$ and $R=I$.

One state-space realization for this system in discrete time is

$$A=\begin{bmatrix}0.8877 & -0.0346\\ 0.1194 & 0.9812\end{bmatrix},\quad B=\begin{bmatrix}-0.0827 & 0.0395\\ 0.0212 & 0.0026\end{bmatrix},$$
$$C=\begin{bmatrix}0 & 1.2472\end{bmatrix}.$$

The system is again (A,B) invariant with respect to the null space of C. In contrast, with the addition of the extra degree of freedom, the closed-loop system with the invariant feedback law

$$K_{\overline{C}}=\begin{bmatrix}-5.5486 & -45.5972\\ -0.6805 & -5.5921\end{bmatrix} \tag{41}$$

is stabilizable. Figure 9 shows a comparison of the closed-loop responses for the constrained regulator with $N=9$ and the unconstrained regulator subject to the initial condition is $x_0=[1,\ -3]^T$. Figure 10 shows the phase portraits of the closed-loop responses for the constrained and unconstrained regulator. Once again, the output response for the constrained regulator appears superior to the unconstrained output response. Notice, however, that the input is far more aggressive for the constrained regulator.

The combined problem of both state and input constraints is solved by reconsidering Eq. 34 after making the following substitutions

$$B\leftarrow B\mathfrak{N}_{\overline{D}},\quad R\leftarrow \mathfrak{N}_{\overline{D}}^T R\,\mathfrak{N}_{\overline{D}},\quad M\leftarrow M\mathfrak{N}_{\overline{D}}. \tag{42}$$

It is not difficult to prove that the proposed control algorithm is asymptotically stable. Convergence of the regulator is straightforward to demonstrate using standard arguments (for example, see Keerthi and Gilbert (1988)). Establishing nominal stability is more subtle than the usual arguments. In particular, the definition of stability needs to be adjusted to account for perturbations only in the feasible region.

Concluding Remarks

The main contribution of this article has been to establish techniques for handling inequality constraints active at steady state, a case that has not been treated in previous model predictive control (MPC) theory. Through a series of examples, we show how this case is significant in applications.

As an alternative to the approach outlined in this article, one could consider moving any inequality constraint line that passes through the origin a small distance away from the origin, after which existing theory would apply. Choosing this distance is problematic, however. If a small distance is chosen, the output admissible set may be small, and the required horizon may be large and the on-line computation is inefficient. If a somewhat larger distance is chosen, the economic

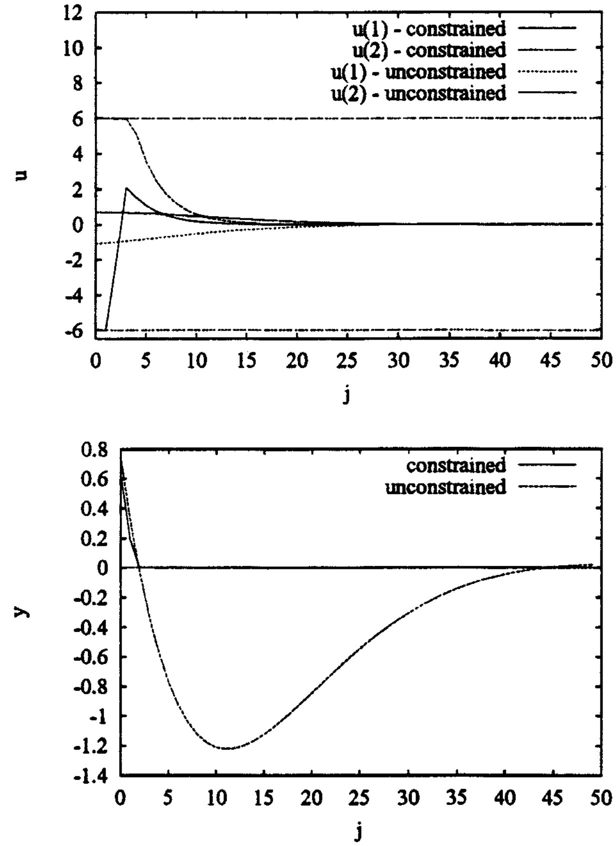

Figure 9. Comparison of closed-loop responses for Example 8.

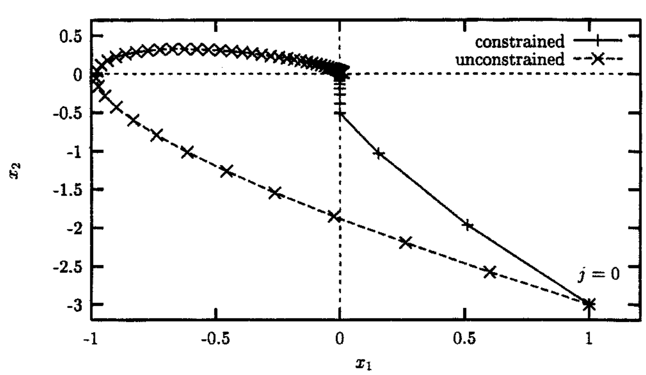

Figure 10. Closed-loop phase portraits for Example 8.

performance of the plant suffers because the steady-state target is no longer close to the true plant constraints. Of course, one could always avoid the issue by using a finite horizon and terminal constraint, but that choice is not as good as the approach outlined here.

Acknowledgments

The authors gratefully acknowledge the financial support of the industrial members of the Texas-Wisconsin Modeling and Control Consortium and NSF through grant No. CTS-9708497. We are grateful to the three referees of the original version of this article, for their insightful comments.

Literature Cited

Berge, C., *Topological Spaces, including a Treatment of Multi-Valued Functions, Vector Spaces and Convexity*, Oliver and Boyd, Edinburgh and London (1963).

Chmielewski, D., and V. Manousiouthakis, "On Constrained Infinite-Time Linear Quadratic Optimal Control," *Sys. Cont. Let.*, **29**, 121 (1996).

de Oliveira, N. M. C., and L. T. Biegler, "Constraint Handling and Stability Properties of Model-Predictive Control," *AIChE J.*, **40**, 1138 (1994).

Fletcher, R., *Practical Methods of Optimization*, Wiley, New York (1987).

Frank, M., and P. Wolfe, "An Algorithm for Quadratic Programming," *Naval Res. Logist. Quart.*, **3**, 95 (1956).

Genceli, H., and M. Nikolaou, "Robust Stability Analysis of Constrained l_1-Norm Model Predictive Control," *AIChE J.*, **39**, 1954 (1993).

Gilbert, E. G., and K. T. Tan, "Linear Systems with State and Control Constraints: The Theory and Application of Maximal Output Admissible Sets," *IEEE Trans. Auto. Cont.*, **36**, 1008 (1991).

Hager, W. W., "Lipschitz Continuity for Constrained Processes," *SIAM J. Cont. Opt.*, **17**, 321 (1979).

Keerthi, S. S., "Optimal Feedback Control of Discrete-Time Systems with State-Control Constraints and General Cost Functions," PhD Thesis, Univ. of Michigan (1986).

Keerthi, S. S., and E. G. Gilbert, "Optimal Infinite-Horizon Feedback Laws for a General Class of Constrained Discrete-Time Systems: Stability and Moving-Horizon Approximations," *J. Optim. Theory Appl.*, **57**, 265 (1988).

Mayne, D. Q., "Optimization in Model Based Control," *Proc. IFAC Symp. Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes*, Helsingor, Denmark, p. 229 (1995).

Michalska, H., and D. Q. Mayne, "Robust Receding Horizon Control of Constrained Nonlinear Systems," *IEEE Trans. Auto. Cont.*, **38**, 1623 (1993).

Muske, K. R., "Steady-State Target Optimization in Linear Model Predictive Control," *Proc. Amer. Control Conf.*, Albuquerque, NM, 3597 (1997).

Muske, K. R., and J. B. Rawlings, "Model Predictive Control with Linear Models," *AIChE J.*, **39**, 262 (1993).

Prett, D. M., and M. Morari, eds., *The Shell Process Control Workshop*, Butterworths, Boston (1987).

Rao, C. V., S. J. Wright, and J. B. Rawlings, "On the Application of Interior Point Methods to Model Predictive Control," *J. Optim. Theory Appl.*, **99**, 723 (1998).

Rawlings, J. B., and K. R. Muske, "Stability of Constrained Receding Horizon Control," *IEEE Trans. Auto. Cont.*, **38**, 1512 (1993).

Ricker, N. L., T. Subrahmanian, and T. Sim, "Case Studies of Model-Predictive Control in Pulp and Paper Production," *Proc. IFAC Workshop on Model Based Process Control*, T. J. McAvoy, Y. Arkun, and E. Zafiriou, eds., Pergamon Press, Oxford, p. 13 (1988).

Rossiter, J. A., M. J. Rice, and B. Kouvaritakis, "A Robust Stable State-Space Approach to Stable Predictive Control Stategies," *Proc. Amer. Control Conf.*, Albuquerque, NM, p. 1640 (1997).

Scokaert, P. O. M., and J. B. Rawlings, "Infinite Horizon Linear Quadratic Control with Constraints," *Proc. IFAC World Cong.*, San Francisco (1996).

Scokaert, P. O., and J. B. Rawlings, "On Infeasibilities in Model Predictive Control," J. C. Kantor, C. E. García, and B. Carnahan, eds., *Chemical Process Control—V*, CACHE, AIChE, p. 331 (1997).

Scoakert, P. O., and J. B. Rawlings, "Constrained Linear Quadratic Regulation," *IEEE Trans. Auto. Cont.*, **43**, 1163 (1998).

Sontag, E. D., *Mathematical Control Theory*, Springer-Verlag, New York (1990).

Sznaier, M., and M. J. Damborg, "Suboptimal Control of Linear Systems with State and Control Inequality Constraints," *Proc. Conf. on Decision and Control*, 761 (1987).

Zheng, A., and M. Morari, "Stability of Model Predictive Control with Mixed Constraints," *IEEE Trans. Auto. Cont.*, **40**, 1818 (1995).

Appendix A: Proof of Uniqueness for Target Calculation

Without loss of generality, we ignore the inequality constraints on the decision varaibles and consider only the constraints given by Eq. 10a.

Theorem 1. If Q_{ss}, $R_{ss} > 0$, $q_{ss} > 0$, and (A,C) is detectable, then the solution to Eq. 9 subject to the constraints given by Eq. 10a is unique.

Proof. Using the Hautus lemma (Sontag, 1990), detectability implies the following rank condition

$$\operatorname{rank}\left(\mathcal{H} \triangleq \begin{bmatrix} \lambda I - A \\ C \\ C \end{bmatrix}\right) = n \tag{A1}$$

for all $\lambda \in \mathbb{C}$ with magnitude greater than or equal to 1. It is sufficient to consider only the extended Hautus matrix $\mathcal{H}$ with $\lambda = 1$. Since rank $\mathcal{H} = n$, $\mathcal{H}$ has full column rank. It then follows that x_{ss} is uniquely determined from the following equation

$$\mathcal{H} x_{ss} = \begin{bmatrix} B(u_{ss} + d) \\ \bar{y} - p - (\eta - t_1) \\ \bar{y} - p + (\eta - t_2) \end{bmatrix} \tag{A2}$$

where t_1 and t_2 are positive slacks for the inequality constraints (Eq. 10a). If $\bar{x}_{ss} \neq x_{ss}$ is another solution, then there necessarily exists another solution $(\bar{u}_{ss}, \bar{\eta} - \bar{t}_{\{1,2\}}) \neq (u_{ss}, \eta - t_{\{1,2\}})$. Since the positive slack accounts for $\eta = |\bar{y} - Cx_{ss} - p|$, η is uniquely determined by $(\eta - t_{\{1,2\}})$. However, since the objective function is a strictly convex function of u_{ss} and η, $\bar{x}_{ss}$ cannot be another solution without contradicting optimality.

Remark 1. An additional consequence of a unique target is that the target calculation is stable to perturbations. Since the quadratic program is continuous in a point-to-set topology, uniqueness of the target guarantees that the solution is continuous in a point-to-point topology (Berge, 1963).

Appendix B: State Constrained Linear Quadratic Regulator

In this section, we describe sufficient conditions for the construction of a state constrained linear feedback controller. The key concept is (A, B) invariance with respect to the null space of $\bar{C}$. For further details of (A, B) invariance with respect to an arbitrary subspace (see Section 4.3 of Sontag, 1990).

Definition 1. The null space of $\bar{C}$ is (A, B) invariant if and only if $\forall w, \exists v$ such that $\bar{C}w = 0$ implies $\bar{C}(Aw + Bv) = 0$.

Theorem 2. The null space of $\bar{C}$ is (A, B) invariant if and only if there exists a linear feedback control law that constrains the evolution of (A, B) to the subspace $\bar{C}w = 0$.

The sufficiency is immediate (Sontag, 1990). Before proving necessity of Theorem 2, we first derive necessary and sufficient conditions for the null space of $\bar{C}$ to be (A, B) invariant. Let $\mathfrak{N}_{\bar{C}}$ be an orthonormal basis for the null-space of $\bar{C}$ and let $\zeta = -\mathfrak{N}_{\bar{C}}^T w$. We recast the constraints (Eq. 32) as

$$\bar{C}A\,\mathfrak{N}_{\bar{C}}\zeta = \bar{C}Bv. \tag{B1}$$

Let $\mathfrak{R}_{(\cdot)}$ denote the range space of $(\cdot)$.

Lemma 1. The null space of $\bar{C}$ is (A, B) invariant if and only if $\mathfrak{R}_{\bar{C}A\mathfrak{N}_C} \subseteq \mathfrak{R}_{\bar{C}B}$.

Proof. Suppose first that $\mathfrak{R}_{\bar{C}A\mathfrak{N}_{\bar{C}}} \subseteq \mathfrak{R}_{\bar{C}B}$. It follows directly that $\forall\zeta, \exists v$ such that $\bar{C}A\,\mathfrak{N}_{\bar{C}}\zeta = \bar{C}Bv$ since the column space of $B\bar{C}$ contains the columns space of $\bar{C}A\,\mathfrak{N}_{\bar{C}}$. Hence, $\bar{C}$ is (A, B) invariant as claimed.

Now, suppose that the null space of $\bar{C}$ is (A, B) invariant. Then $\forall\zeta, \exists v$ such that $\bar{C}A\,\mathfrak{N}_{\bar{C}}\zeta = \bar{C}Bv$. Therefore, by definition, $\mathfrak{R}_{\bar{C}A\mathfrak{N}_{\bar{C}}} \subseteq \mathfrak{R}_{\bar{C}B}$ as claimed.

Let O_A and O_B denote the orthonormal bases for the columns spaces of $\bar{C}A\,\mathfrak{N}_{\bar{C}}$ and $\bar{C}B$ respectively.

Corollary 1. The null space of $\bar{C}$ is (A, B) invariant if and only if $(O_B O_B^T - I)O_A = 0$.

Proof. An equivalent condition for $\mathfrak{R}_{\bar{C}A\mathfrak{N}_{\bar{C}}} \subseteq \mathfrak{R}_{\bar{C}B}$ is that $\forall\zeta, \exists v$ such that $O_A\zeta = O_B v$. Solving for v yields $v = O_B^T O_A \zeta$. Direct substitution yields the desired result.

Proof of Theorem 2. We construct a feedback law that constrains w to the null-space of $\bar{C}$ by decomposing the operator $\bar{C}B$ into its range space and null spaces yielding $v = -(\bar{C}B)^+ CAw + \mathfrak{N}_{\bar{C}B}\, p^{\bar{C}}$, where $p^{\bar{C}}$ is the input projected to the null space of $\bar{C}B$ and $(\cdot)^+$ denotes the pseudo-inverse. If $K_{\bar{C}} \triangleq -(\bar{C}B)^+ CA$, then we construct the linear feedback law $v = Kw + \mathfrak{N}_{\bar{C}B}\, p^{\bar{C}}$ as claimed.

Remark 2. The vector $p^{\bar{C}}$ represents the excess degrees of freedom in the inputs with respect to the constraint $\bar{C}w = 0$. Since the control law constrains the system to the subspace $\bar{C}w = 0$ for all $p^{\bar{C}}$, we construct the optimal state constrained regulator by first constructing a linear quadratic regulator K for the system $(A + BK_{\bar{C}}, B\mathfrak{N}_{\bar{C}B})$. The full state constrained regulator is obtained by combining the two linear regulators as follows

$$v = \left(K_{\bar{C}} + \mathfrak{N}_{\bar{C}B} K\right) w. \tag{B2}$$

Manuscript received Aug. 17, 1998, and revision received Mar. 26, 1999.